

Statistics for Radiologists

Harold L. Kundel, MD, *Editor*

Comparison of Receiver Operating Characteristic Curves on the Basis of Optimal Operating Points

Ethan J. Halpern, MD¹, Michael Albert, PhD¹, Abba M. Krieger, PhD², Charles E. Metz, PhD³, Andrew D. Maidment, PhD¹

Rationale and Objectives. We developed a method of comparing receiver operating characteristic (ROC) curves on the basis of the utilities associated with their optimal operating points (OOPs).

Methods. OOPs were computed for paired ROC curves on the basis of isocost lines in ROC space with slopes ranging from 0.1 to 3.0. For each pair of OOPs corresponding to a single isocost slope, the difference in costs and the variance of this difference was computed. A sensitivity analysis was thus obtained for the difference between the two curves over a range of isocost slopes. Three published data sets were evaluated using this technique, as well as by comparisons of areas under the curves and of true-positive fractions at fixed false-positive fractions.

Results. The OOPs of paired ROC curves often occur at different false-positive fractions. Comparisons of ROC curves on the basis of OOPs may provide results that differ from comparisons of curves at a fixed false-positive fraction.

Conclusion. ROC curves may be compared on the basis of utilities associated with their OOPs. This comparison of the optimal performance of two diagnostic tests may differ from conventional statistical comparisons.

Key Words. Receiver operating characteristic curve; area under the curve; optimal operating point; statistical comparison.

Diagnostic tests often provide a continuous value that may be interpreted as a dichotomous result (normal or abnormal). The true-positive and false-positive

rates of the dichotomous interpretation depend on the underlying distributions of test results (Fig. 1A) in the normal and abnormal populations, as well as on the cutoff value used to discriminate between normal and abnormal populations. As the cutoff value is varied, a receiver operating characteristic (ROC) curve is generated (Fig. 1B). For any given clinical scenario, there is an optimal operating point (OOP) on the ROC curve that defines the most appropriate cutoff value to discriminate a positive from a negative test result.

In terms of cost-benefit analysis, the OOP on a ROC curve maximizes the expected utility of a diagnostic test. The utility of a diagnostic test depends on the prior expectation of disease (or disease prevalence) and the relative costs incurred by a false-positive or a false-negative result. The slope of a line of "isoutilility" in ROC space is given by

$$\frac{(\text{prevalence of disease})}{1 - (\text{prevalence of disease})} \times \frac{(\text{cost of false-positive result})}{(\text{cost of false-negative result})} \quad (1)$$

This slope defines a family of parallel lines. The OOP on a ROC curve must be tangent to the highest line of isoutilility that intersects with the ROC curve. The slope of the ROC curve at its OOP will be equal to the slope of isoutilility [1].

By analogy with laboratory tests that provide a continuous numeric result, imaging studies provide a result (supporting or refuting a particular diagnosis) with a variable confidence level. The ROC curve for a diagnostic imaging study plots the true- and false-positive rates

From the ¹Department of Radiology, Jefferson Medical College of Thomas Jefferson University Hospital, Philadelphia, PA; ²Department of Statistics, Wharton School, University of Pennsylvania, Philadelphia, PA; and ³Department of Radiology, University of Chicago Medical Center, Chicago, IL.

This work was partly supported by a grant from General Electric.

Ethan J. Halpern, MD, is a member of the General Electric-Association of University Radiologists' Radiology Research Academic Fellowship Program.

Address reprint requests to E. J. Halpern, MD, Department of Radiology, Thomas Jefferson University, 132 S. 10th St., Philadelphia, PA 19107-5244.

Received August 11, 1995, and accepted for publication after revision November 20, 1995.

Acad Radiol 1996;3:245-253

© 1996, Association of University Radiologists

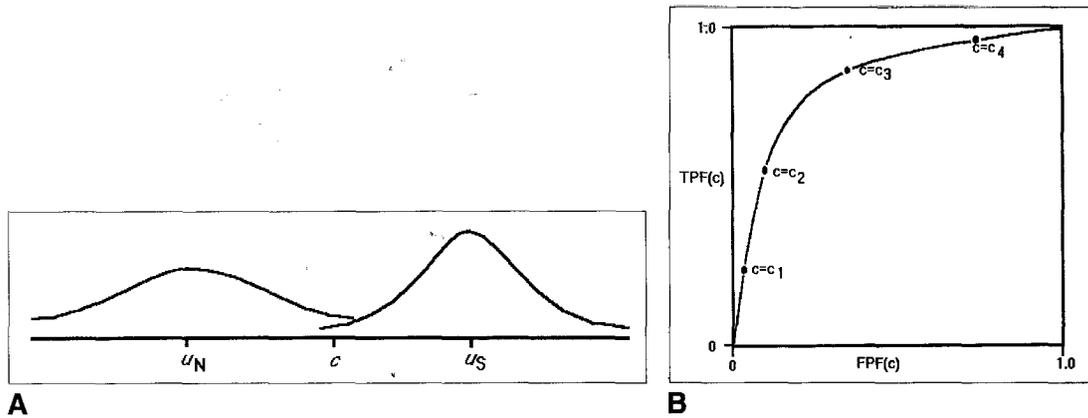


FIGURE 1. **A**, Distribution of signal and noise from which the classic receiver operating characteristic (ROC) curve is derived. The dichotomous interpretation of the diagnostic study is positive when the value obtained is greater than *c*. **B**, In the resulting ROC curve, each value of *c* defines an expected true-positive fraction (TPF) and an expected false-positive fraction (FPF). N = normal, S = abnormal.

of the study at these various confidence levels. The OOP for a diagnostic imaging study defines the confidence level that will provide the best test performance from the cost-benefit analysis perspective.

ROC curves are used both to evaluate individual diagnostic tests and to compare the relative accuracy of competing diagnostic tests. The area under the ROC curve (AUC) provides a summary index to evaluate a diagnostic test [2] and may be used to provide a statistical comparison of competing diagnostic tests [2, 3]. The AUC evaluates the accuracy of a diagnostic test over the full range of possible discriminating cutoff values. In practice, however, most tests are (or at least should be) applied with a discriminating value close to the OOP. Thus, the AUC may not be truly representative of the diagnostic accuracy of a test as it is used in clinical practice.

To provide a comparison of the more clinically relevant portions of ROC curves, a method has been described to compare true-positive fractions (TPFs) at preselected false-positive fractions (FPFs) [4]. In this technique, it is assumed that two diagnostic tests should be compared with their respective discriminating values adjusted to achieve identical false-positive rates. Often, however, the OOP of one test is at a different false-positive level than the OOP of a competing diagnostic study. Under such circumstances, a comparison of TPFs at a preselected FPF does not properly compare the optimal utilization of both tests.

It has been suggested recently that a method should be developed for comparing ROC curves on the basis of the cost (or utility) associated with the OOP for each diagnostic study [5]. We developed such a technique and applied this technique to three data sets obtained from the radiology literature [6-8]. These data sets were

analyzed previously by comparing AUCs or TPFs at set FPFs. In this article, we analyze these data sets on the basis of the costs associated with the OOPs and compare our results with the conventional techniques used in the original publications.

MATERIALS AND METHODS

Determination of the OOP on a ROC Curve

Assume that we have a diagnostic study that must distinguish between noise and signal. The magnitude of both the noise and the true signal are normally distributed with means μ_N and μ_S , where N denotes normal and S abnormal. As detailed in the Appendix, a ROC curve for this scenario is defined by two parameters, *a* and *b*. Parameter *a* represents the normalized difference between the means: $a = (\mu_S - \mu_N)/\sigma_S$. Parameter *b* represents the ratio of the standard deviations: $b = \sigma_N/\sigma_S$. Equation 1 provides the slope of a line of isoutility in ROC space. This line will be tangent to a ROC curve at its OOP. We call this slope β . Assuming that we know the value of β for a particular diagnostic situation, we need to find the OOP on the ROC curve for a diagnostic test. The FPFs and TPFs at the OOP are given by

$$1) \text{ for } b=1, \text{ FPF}_{\text{OOP}}(a,\beta) = \Phi \left[\frac{-a/2 - \ln(\beta)/a}{b} \right] \tag{2}$$

$$\text{TPF}_{\text{OOP}}(a,\beta) = \Phi \left[\frac{+a/2 - \ln(\beta)/a}{b} \right]$$

$$2) \text{ for } b \neq 1, \text{ FPF}_{\text{OOP}}(a,b,\beta) = \Phi \left\{ \frac{[ab - \sqrt{a^2 + 2(1-b^2)\ln(\beta/b)}]/(1-b^2)}{b} \right\}$$

$$\text{TPF}_{\text{OOP}}(a,b,\beta) = \Phi \left\{ \frac{[a - b\sqrt{a^2 + 2(1-b^2)\ln(\beta/b)}]/(1-b^2)}{b} \right\}$$

where Φ is the area under the standard normal curve to the left of the value within the parentheses.

Computation of the Cost of a Diagnostic Test

The slope of the isoutility line in ROC space, β , defines the relative utility of true- and false-positive results for a given prevalence of disease. At any point on a ROC curve, the expected cost of a diagnostic test is determined by the FPF and TPF and is given by

$$K = \lambda[\beta(FPF) - (TPF)] + (C_{\text{study}}), \quad (3)$$

where K is the expected cost (negative utility), λ represents a constant that translates K into the appropriate units of cost, and C_{study} represents the intrinsic costs of the diagnostic study, including actual monetary costs as well as potential morbidity–mortality that may result from the test.

Comparison of the Difference in Cost Between Two Diagnostic Studies

In our derivation, assume that C_{study} is similar for competing diagnostic studies (or that it differs by an amount that is not significant in the overall treatment plan for the patient). Hence, the difference in utility between the two competing modalities depends only on their respective TPF and FPF, on λ , and on β . Because the two tests are applied to the same diagnostic situation, the values of λ and β are identical for the cost functions of each modality.

The variance of K cannot be expressed in closed form. However, fluctuation of K near the OOP can be approximated as a linear function of small fluctuations in a and b (the two parameters that define the ROC curve). The partial derivatives of K with respect to a and b at the OOP are given by

$$\begin{aligned} \partial K / \partial a &= -(\lambda / \sqrt{2\pi}) \exp(-0.5Y_0^2) \\ \partial K / \partial b &= (\lambda / \sqrt{2\pi}) X_0 \exp(-0.5Y_0^2), \end{aligned} \quad (4)$$

where X_0 represents the normal deviate value that corresponds to $(1 - FPF)$ at the OOP, and Y_0 represents the normal deviate value that corresponds to $(1 - TPF)$ there. Thus, if the variances and covariance of a and b are known at the time a ROC curve is constructed, the variance of K may be approximated from the variances and covariance of its linear components.

For any two diagnostic tests, the difference in utility between the OOPs of the two tests ($K_1 - K_2$) is calcu-

lated from equation 3. The variance of this difference is approximated from the variances and covariance of the a and b parameters of each ROC curve (Appendix). A Z statistic may be calculated as the ratio of $(K_1 - K_2)$ to the standard deviation of $(K_1 - K_2)$. The constant λ is eliminated because it appears in both the cost and its standard deviation. As shown in the Appendix, the resulting Z statistic depends only on a , b , and β .

When a ROC curve is determined by the maximum-likelihood technique, the values of the curve parameters a and b , and the variances and covariance of these parameters, are estimated from ordinal image-reading or test-result data. In general, the value of β for any diagnostic test is not known with certainty because it depends on disease prevalence and the perceived cost of underdiagnosis and overdiagnosis (equation 1). Given an estimate of β , we may calculate the Z statistic for a comparison of utilities at the OOPs of two competing diagnostic modalities. To allow for uncertainty in the true value of β , however, we calculated the Z statistic over a range of possible values. The result of this calculation is a sensitivity analysis for the difference in utilities as a function of β .

Comparison of Data Sets

In the radiology literature, we found three data sets [6–8] describing competing diagnostic modalities that were previously studied by conventional ROC analysis. Each of these data sets was analyzed previously on the basis of the AUCs or on the basis of calculated TPFs at set FPFs. In addition to repeating these analyses, we calculated the costs associated with the OOPs of the respective ROC curves. For the comparison of optimal operating costs, the value of β was varied from 0.1 to 3.0.

RESULTS

Table 1 shows the results of a comparison of gallium citrate imaging at the Peter Bent Brigham Hospital (PBBH; Boston, MA) using an Anger camera with imaging at Johns Hopkins Hospital (JHH; Baltimore, MD) using a rectilinear scanner for the diagnosis of a focal source of sepsis [6]. In accordance with the published results, we determined that the AUCs were 0.767 ± 0.065 (mean \pm standard error) at the PBBH and 0.680 ± 0.077 at the JHH. A two-tailed Z test of the differences in these areas was not significant. A comparison between PBBH and JHH of TPFs from these ROC curves over a range of FPFs from 0.05 to 0.95 also was not significant. Likewise, a comparison of utility associated with the OOPs over a

TABLE 1: ROC Analysis for Gallium Citrate Imaging at the Peter Bent Brigham Hospital (PBBH) Using an Anger Camera and at Johns Hopkins Hospital (JHH) Using a Rectilinear Scanner

Variable	PBBH	JHH	<i>p</i>			
ROC parameters	<i>a</i> = 0.8177 <i>b</i> = 0.5039 Area = 0.767	<i>a</i> = 0.5843 <i>b</i> = 0.7449 Area = 0.680	— .4			
	FPF	TPF for PBBH	TPF for JHH			
	0.1	0.57	0.36	.25		
	0.2	0.65	0.48	.25		
	0.3	0.71	0.58	.31		
	0.4	0.76	0.65	.41		
	0.5	0.79	0.72	.56		
	0.6	0.83	0.78	.70		
	0.7	0.86	0.84	.84		
	0.8	0.89	0.89	.96		
	0.9	0.93	0.94	.92		
Isocost slope β	OOP PBBH		OOP JHH			
	FPF	TPF	FPF	TPF		
	0.1	1.0	1.0	1.0	—	
	0.5	0.30	0.71	1.0	.45	
	1.0	0.12	0.59	0.22	0.51	.23
	1.5	0.07	0.53	0.11	0.37	.23
	2.0	0.05	0.49	0.07	0.30	.24
	2.5	0.04	0.47	0.05	0.25	.25
	3.0	0.03	0.45	0.03	0.21	.25

The ROC parameters were calculated from the raw data; they differ slightly from those provided by McNeil and Hanley [4]. FPF = false-positive fraction, TPF = true-positive fraction, OOP = optimal operating point.

TABLE 2: ROC Analysis for Interpretation of Computed Tomography Studies of the Head with and Without a Clinical History

Variable	With History	Without History	<i>p</i>			
ROC parameters	<i>a</i> = 3.6 <i>b</i> = 1.29 Area = 0.986	<i>a</i> = 1.8 <i>b</i> = 0.59 Area = 0.939	.06			
	FPF	TPF with history	TPF without history			
	0.1	0.97	0.85	.03		
	0.2	0.99	0.91	.03		
	0.3	1.00	0.93	.05		
	0.4	1.00	0.95	.06		
	0.5	1.00	0.96	.08		
	0.6	1.00	0.97	.09		
	0.7	1.00	0.98	.10		
	0.8	1.00	0.99	.12		
	0.9	1.00	1.00	.13		
Isocost slope β	OOP with history		OOP without history			
	FPF	TPF	FPF	TPF		
	0.1	0.18	0.99	0.56	0.97	.02
	0.5	0.09	0.97	0.15	0.88	.01
	1.0	0.07	0.95	0.08	0.83	.02
	1.5	0.05	0.94	0.05	0.80	.05
	2.0	0.05	0.92	0.04	0.78	.08
	2.5	0.04	0.91	0.03	0.76	.13
	3.0	0.04	0.90	0.03	0.75	.18

ROC = receiver operating characteristic, FPF = false-positive fraction, TPF = true-positive fraction, OOP = optimal operating point.

range of 0.1–3.0 was not significant. In this data set, all three methods led to the same conclusion.

Table 2 shows the results of a comparison for reading computed tomography studies of the head with and

without a clinical history [7]. The AUCs were 0.986 ± 0.009 with a clinical history and 0.939 ± 0.029 without a clinical history. The difference between these areas was marginally significant (two-tailed *Z* test, $p = .06$). A

comparison of TPFs from these ROC curves was significant ($p < .05$) for FPFs of 0.10–0.30. A comparison of utilities associated with the OOPs was significant for values of β ranging from 0.1 to 1.5. This range corresponded to FPFs of 0.05–0.18 for interpretations with a clinical history and 0.05–0.56 for interpretations without a clinical history. The results of all three analyses suggested a significant difference between readings with and without a clinical history. However, the comparison based on OOPs showed a significant difference between the curves over a larger range of false-positive values than was found in comparisons at fixed FPFs. The comparison based on OOPs did not necessarily compare points at identical FPFs. For example, at $\beta = 0.1$, the cost of interpretation with a clinical history at an FPF of 0.18 was compared with the cost of interpretation without a clinical history at an FPF of 0.56.

Table 3 shows a comparison of two Doppler ultrasound measurements for the detection of renal artery stenosis: Doppler interrogation of the main renal artery (the renal artery-to-aortic ratio [RAR]) and Doppler interrogation of segmental vessels (minimum early systolic acceleration [ESA]) [8]. The AUCs were 0.798 ± 0.070 for RAR and 0.926 ± 0.044 for ESA. The difference between these areas was marginally significant (two-

tailed Z test, $p = .05$). A comparison of TPFs for these two techniques showed a significant difference at FPFs of 0.05–0.30. A comparison of the two techniques based on OOPs indicated that ESA was significantly superior to RAR over a range of $\beta = 0.3$ –3.0. This range corresponded to a range of FPFs of 0.03–0.70 for RAR. In this situation, the comparison based on OOPs provided evidence that ESA is superior to RAR over the entire range of clinically feasible FPFs for RAR.

DISCUSSION

The most established technique for comparing ROC curves is that of the AUC [2, 3]. The AUC represents an average of the true-positive rate over the entire range of possible false-positive rates and is inversely related to the average of the false-positive rate over the full range of true-positive rates. Theoretically, the area may vary from 0.5 to 1.0. Area provides a single number to characterize the ROC curve, thereby simplifying the comparison of any two curves. However, in any real clinical situation, only a small portion of the ROC curve (near the OOP) should be used when interpreting a diagnostic study result. Thus, the comparison of two ROC curves on the basis of their respective areas will include portions of the ROC curves that are not relevant in clinical practice.

TABLE 3: ROC Analysis for Doppler Detection of Renal Artery Stenosis by Renal Artery-to-Aortic Ratios (RARs) in the Main Renal Artery and by Early Systolic Acceleration (ESA) in Segmental Vessels

Variable	RAR		ESA		<i>p</i>
ROC parameters	<i>a</i> = 0.9745 <i>b</i> = 0.6001 Area = 0.798		<i>a</i> = 1.5840 <i>b</i> = 0.4421 Area = 0.926		.05
	FPF	TPF with RAR	TPF with ESA		
	0.1	0.58	0.85		.03
	0.2	0.68	0.89		.03
	0.3	0.75	0.91		.05
	0.4	0.79	0.93		.08
	0.5	0.84	0.94		.13
	0.6	0.87	0.96		.19
	0.7	0.90	0.97		.28
	0.8	0.93	0.97		.38
	0.9	0.96	0.98		.52
Isocost slope β	OOP with RAR		OOP with ESA		
	FPF	TPF	FPF	TPF	
	0.1	1.0	0.67	0.96	.90
	0.5	0.34	0.12	0.85	.02
	1.0	0.14	0.06	0.81	.02
	1.5	0.09	0.04	0.79	.03
	2.0	0.06	0.03	0.77	.03
	2.5	0.05	0.02	0.76	.03
	3.0	0.03	0.02	0.75	.03

ROC = receiver operating characteristic, FPF = false-positive fraction, TPF = true-positive fraction, OOP = optimal operating point.

To achieve a more clinically meaningful comparison of ROC curves, techniques have been described to focus the comparison on limited portions of the curves [4]. In these techniques, it is assumed that two ROC curves should be compared at similar TPFs or FPFs. However, as demonstrated in Table 1 for $\beta = 0.5$ – 2.5 , in Table 2 for $\beta = 0.1$ – 1.0 , and in Table 3 for $\beta = 0.1$ – 3.0 , the OOPs on two ROC curves may occur in different portions of the curves. Furthermore, when comparing partial areas, one encounters the additional uncertainty of whether to compare the areas below or to the right of the curve segments to be compared [9]. We have therefore developed a method for comparing ROC curves that is based on the utilities associated with OOPs. The OOP of a ROC curve corresponds to the discriminating cutoff (between normal and abnormal) that is most appropriate to use when the test is applied in clinical practice.

The utility associated with the OOP of a ROC curve is a numeric estimate of the expected value of the clinical outcomes associated with that curve. Utility, as assessed by a standard reference gamble, is a measure of preference that incorporates patients' attitudes about various states of health and disease [10]. Unfortunately, utilities often vary among individuals as well as within one individual over time. Nonetheless, utility assessment provides an important quantitative tool for medical decision making. When faced with a choice between two competing diagnostic modalities represented by two ROC curves, the ROC curve with the greater utility should be chosen to maximize patients' well-being.

The location of the OOP of a ROC curve depends on the expected utilities and costs associated with true-positive and false-positive test results, the prevalence of the disease, and the shape of the ROC curve. The first two of these three factors are summarized in the variable β (equation 1). The shape of the ROC curve is defined by the parameters a and b . For any particular clinical situation, the value of β must be estimated in order to calculate the OOP. Because this estimate depends on an individual's evaluation of the relative utilities of different states of health and disease, the estimate may be considered subjective and will be subject to criticism. To overcome this issue, we propose that the comparison of utilities be performed over a range of OOPs. Ideally, the range of clinically relevant values of β should be defined before the statistical comparison is performed.

We have found empirically that by varying the value of β from 0.1 to 3.0, the comparison of OOPs tends to

include portions of the ROC curves that we would consider clinically relevant. When the value of β is less than 0.1, the OOP occurs along the upper portion of the ROC curve and is usually so far to the right that the test result is almost always interpreted to be negative. When the value of β is greater than 3.0, the OOP occurs along the lower portion of the ROC curve and is usually so far to the left that the test result is almost always interpreted to be positive. These extreme values of β correspond to clinical situations in which a diagnostic test probably is not relevant. Thus, in general, an appropriate comparison of ROC curves may be performed as a sensitivity analysis that includes a subset of OOPs over the range of 0.1–3.0.

In our technique for comparing two ROC curves, it is assumed that the costs of the diagnostic studies themselves, C_{study} , are relatively similar and may be disregarded. For many clinical situations, this assumption holds true. However, when the cost of the diagnostic study is high relative to the potential cost of the disease under study, or when the diagnostic study involves potential risk to the patient, the cost of the study must be considered. Unfortunately, this cost introduces new unknowns into the C_{study} term in equation 3 and into the calculation of variance for the cost difference. If an estimate of ΔC_{study} is included in the calculation of the expected cost difference, the λ term (in equation 3) must be known in order to calculate the Z statistic. Thus, in situations in which such a limitation exists, it may be more appropriate to use one of the more traditional methods of comparison (i.e., area or TPF at a fixed FPF) rather than to guess at the OOP.

The OOP approach to ROC curve comparison forces the analysis to consider the clinically relevant portions of the ROC curves. The clinically relevant portions of two paired ROC curves may occur at similar or different FPFs. The benefit of our method can be clearly demonstrated with an example. In Table 2, when $\beta = 1.5$, the clinically relevant comparison occurs at an FPF of 0.05 for both types of interpretations. A simple comparison of TPFs is appropriate in this scenario. However, when $\beta = 0.2$, the clinically relevant comparison occurs at an FPF of 0.14 for interpretations with a clinical history and at an FPF of 0.33 for interpretation without a clinical history. A comparison of TPFs at these different FPFs is meaningless. However, analysis of the utilities associated with these two OOPs does allow a direct comparison of the two ROC curves.

When two diagnostic studies involve comparable costs and risks to the patient, a ROC comparison based

on the OOPs directly compares the most clinically relevant portions of the ROC curves. This type of comparison should provide more clinically meaningful results than a comparison of total AUCs or of TPFs at fixed FPFs. To facilitate this type of analysis, a computer program is available from the authors to compare ROC curves on the basis of the utilities associated with the OOPs as a sensitivity analysis over any selected range of values for β .

REFERENCES

1. Metz CE. Basic principles of ROC analysis. *Sem Nuc Med* **1978**;8:4:283-298.
2. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* **1982**;143:29-36.
3. Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology* **1983**;148:839-843.
4. McNeil BJ, Hanley JA. Statistical approaches to the analysis of receiver operating characteristic (ROC) curves. *Med Decis Making* **1984**;4:2:137-150.
5. Sunshine J. Contributed comment. *Acad Radiol* **1995**;2[suppl]:S72-S74.
6. McNeil BJ, Sanders R, Alderson PO, et al. A prospective study of computed tomography, ultrasound, and gallium imaging in patients with fever. *Radiology* **1981**;139:647-653.
7. McNeil BJ, Hanley JA, Funkenstein HH, Wallman J. The use of paired ROC curves in studying the impact of history on radiography interpretation: CT of the head as a case study. *Radiology* **1983**;149:75-77.
8. Halpern EJ, Needleman L, Nack TL, East SA. Renal artery stenosis: should we study the main renal artery or segmental vessels? *Radiology* **1995**;195:799-804.
9. McClish DE. Analyzing a portion of the ROC curve. *Invest Radiol* **1989**;9:190-195.
10. Sox HC, Blatt MA, Higgins MC, Marton KI. *Medical decision making*. Boston: Butterworth-Heinemann, **1988**:168-169.
11. Metz CE. Some practical issues of experimental design and data analysis in radiological ROC studies. *Invest Radiol* **1989**;24:234-245.
12. Metz CE, Wang P-L, Kronman HB. A new approach for testing the significance of differences between ROC curves measured from correlated data. In: Deconinck F, ed. *Information processing in medical imaging*. Dordrecht, The Hague: Martinus Nijhoff, **1984**:432-445.

APPENDIX

Determination of the Optimal Operating Point

Figure 1 demonstrates the classical situation in which receiver operating characteristic (ROC) analysis is applied. Both the signal and noise in a system are normally distributed. The noise is distributed about a mean of μ_N (N = normal) with a standard deviation of σ_N . The signal is distributed about a mean of μ_S (S = abnormal) with a standard deviation of σ_S . A discriminating value, c , was chosen to distinguish signal from noise. As the value of c is varied, the discriminating power of the test changes. Small values of c will result in large true-positive and false-positive rates. Large values of c will result in small true-positive and false-positive rates. The collection of true- and false-positive rates associated with various values of c define the ROC curve.

Any value of c may be represented as a Z value in either the noise or signal distribution: $Z_N = (c - \mu_N)/\sigma_N$ or $Z_S = (c - \mu_S)/\sigma_S$. Solving both equations for c , we find that

$$\begin{aligned} Z_N \sigma_N + \mu_N &= c = Z_S \sigma_S + \mu_S \\ Z_S &= (\sigma_N/\sigma_S)Z_N - (\mu_S - \mu_N)/\sigma_S \\ Z_S &= bZ_N - a, \end{aligned}$$

which is in the form $Y = bX - a$. Thus, for fixed values of a and b , the Z value corresponding to c in the signal distribution (Y) is a linear function of the Z value corresponding to c in the noise distribution (X). The points on a ROC curve (false-positive fraction [FPF], true-positive fraction [TPF]) may be expressed as a function of c :

$$\begin{aligned} \text{(FPF, TPF)} &= \left[(1/\sqrt{2\pi}) \int_{t=Z_N(c)}^{\infty} \exp(-t^2/2) dt, \right. \\ &\quad \left. (1/\sqrt{2\pi}) \int_{t=Z_S(c)}^{\infty} \exp(-t^2/2) dt \right] \end{aligned}$$

or as a function of the Z values associated with c

$$\begin{aligned} \text{(FPF, TPF)} &= \left[(1/\sqrt{2\pi}) \int_{t=X}^{\infty} \exp(-t^2/2) dt, \right. \\ &\quad \left. (1/\sqrt{2\pi}) \int_{t=Y(X)}^{\infty} \exp(-t^2/2) dt \right] . \end{aligned}$$

Differentiating the points on the ROC curve with respect to X yields

$$\begin{aligned} (\partial \text{FPF}/\partial X, \partial \text{TPF}/\partial X) &= \\ \left[(-1/\sqrt{2\pi}) \exp(-X^2/2), (-b/\sqrt{2\pi}) \exp(-Y^2/2) \right] . \end{aligned}$$

Thus, the slope of the ROC curve ($\partial \text{TPF}/\partial \text{FPF}$) is simply

$$\begin{aligned} [b \exp(-Y^2/2)] / [\exp(-X^2/2)] &= \\ b \exp\{-0.5[(b^2 - 1)X^2 - 2abX + a^2]\} . \end{aligned}$$

The shape of the ROC curve as determined from this equation is shown in Figure 2. When b is greater than one, the X^2 term is dominant and positive; as X becomes a large positive or negative number the slope of the ROC curve approaches zero (Fig. 2A). When b is less than one, the X^2 term is dominant and negative; as X becomes a large positive or negative number, the slope of the ROC curve approaches infinity (Fig. 2B). When $b = 1$, the quadratic term falls out. Then, as X becomes a large positive number (the left side of the ROC curve), the slope approaches infinity; as X becomes a large negative number (the right side of the ROC curve), the slope approaches zero (Fig. 2C).

Isocost lines in ROC space are defined by a positive slope β , which depends on the prevalence of disease and on the relative costs of false-positive and false-negative diagnoses. For any value of β , we can define a family of parallel isocost lines. The optimal operating point of a ROC curve must be tangent to the isocost line. This point can be found by setting the slope of the ROC curve equal to β .

$$\beta = b \exp\{-0.5[(b^2 - 1)X^2 - 2abX + a^2]\}.$$

When b is not equal to one, solving the resulting quadratic equation yields

$$X = [ab \pm \sqrt{a^2 - 2(b^2 - 1) \ln(\beta/b)}] / (b^2 - 1).$$

When b is equal to one, solving the resulting linear equation yields

$$X = a/2 + \ln(\beta)/a.$$

When b is not equal to one, there are two possible solutions to the quadratic. These solutions correspond to two points on the ROC curve that are tangent to an isocost line. Of these two solutions, the isocost line that is higher and to the left in ROC space has greater utility. When b is greater than one, the optimal solution can be found to the right of the point of inflection of the ROC curve (Fig. 3A). When b is less than one, the optimal solution can be found to the left of the point of inflection of the ROC curve (Fig. 3B). Both of these solutions correspond to the value

$$X = [ab - \sqrt{a^2 - 2(b^2 - 1) \ln(\beta/b)}] / (b^2 - 1).$$

X represents the Z value in the noise distribution of the optimal discriminating cutoff. The corresponding Z value in the signal distribution is given by $Y = bX - a$. When b is not equal to one,

$$Y = [a - b \sqrt{a^2 + 2(b^2 - 1) \ln(\beta/b)}] / (b^2 - 1).$$

When $b = 1$, $Y = \ln(\beta)/a - a/2$. The corresponding FPFs and TPFs may be calculated from the areas under the normal curve to the right of the Z values X and Y , respectively. Alternatively, the FPFs and TPFs may be expressed as the areas under the normal curve to the left of $-X$ and $-Y$, respectively. This is the form presented earlier in equation 2.

The Cost Function and Its Variance

Assuming that the intrinsic cost of the diagnostic study, C_{study} , is insignificant relative to the clinical outcome or that the costs of two diagnostic studies that are to be compared are roughly equivalent, the cost function was defined in equation 3 as follows: $K = \lambda[\beta(FPF) - (TPF)]$.

In terms of the values of X and Y at the optimal operating point (call them X_0 and Y_0),

$$K = \lambda \left[\beta \left(\frac{1}{\sqrt{2\pi}} \int_{t=X_0}^{\infty} \exp(-t^2/2) dt \right) - \left(\frac{1}{\sqrt{2\pi}} \int_{t=Y_0}^{\infty} \exp(-t^2/2) dt \right) \right].$$

Differentiating with respect to a

$$\frac{\partial K}{\partial a} = \lambda \left(-\beta \left(\frac{1}{\sqrt{2\pi}} \right) \exp(-X_0^2/2) \frac{\partial X_0}{\partial a} + \left(\frac{1}{\sqrt{2\pi}} \right) \exp(-Y_0^2/2) (b \frac{\partial X_0}{\partial a} - 1) \right).$$

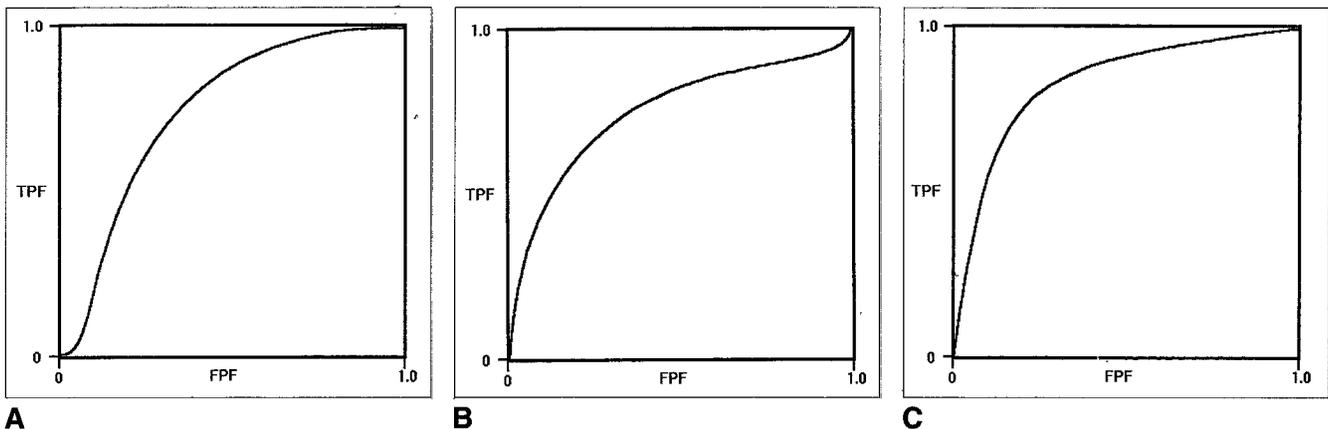
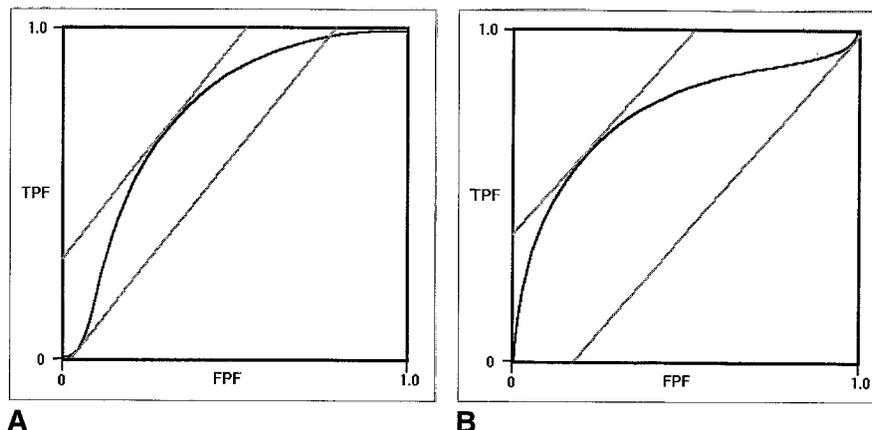


FIGURE 2. A, Receiver operating characteristic (ROC) curve shape. When b is greater than 1, the slope of the ROC curve approaches 0 as the false-positive fraction (FPF) approaches 0 or 1. B, ROC curve shape. When b is less than 1, the slope of the ROC curve approaches infinity as the FPF approaches 0 or 1. C, ROC curve shape. When $b = 1$, the quadratic term disappears. When the FPF approaches 0, the slope approaches infinity. When the FPF is near 1, the slope approaches 0. There is no point of inflection in this curve. TPF = true-positive fraction.

FIGURE 3. A, Isocost lines in receiver operating characteristic (ROC) space. When b is greater than 1, the optimal solution is located where the isocost line is tangent with the ROC curve to the right of the point of inflection on the ROC curve. **B,** Isocost lines in ROC space. When b is less than 1, the optimal solution is located where the isocost line is tangent with the ROC curve to the left of the point of inflection on the ROC curve. TPT = true-positive fraction, FPF = false-positive fraction.



However, at the optimal operating point, we know that $\beta = [b \exp(-Y^2/2)]/[\exp(-X^2/2)]$. Using this equality, we can simplify the partial derivative to

$$\partial K/\partial a = -(\lambda \sqrt{2\pi}) \exp(-0.5 Y_0^2) .$$

Similarly, differentiating K with respect to b ,

$$\partial K/\partial b = \lambda \left[-\beta(1/\sqrt{2\pi}) \exp(-X_0^2/2) \partial X_0/\partial b + (1/\sqrt{2\pi}) \exp(-Y_0^2/2) (b \partial X_0/\partial b + X_0) \right] .$$

Once again, at the optimal operating point, we know that $\beta = [b \exp(-Y^2/2)]/[\exp(-X^2/2)]$. Thus, the derivative simplifies to

$$\partial K/\partial b = -(\lambda \sqrt{2\pi}) X_0 \exp(-0.5 Y_0^2) .$$

For two diagnostic procedures, the difference in costs is simply $K_1 - K_2$. A first-order approximation for the changes in each K_1 and K_2 around their respective optimal operating points may be expressed as a linear function of $\partial K/\partial a$ and $\partial K/\partial b$. Because the variances of a and b are calculated when the ROC curve is constructed, the variance of $K_1 - K_2$ may be expressed as a linear combination of the variances and covariance of a and b for the two curves. Specifically, $\text{variance}(K_1 - K_2) = A^T V A$, where $A^T = (\partial K_1/\partial a_1, \partial K_1/\partial b_1, -\partial K_2/\partial a_2, -\partial K_2/\partial b_2)$ and V represents the covariance matrix, which is provided by Metz's ROCFIT [11] and CORROC [12] algorithms, for example. The Z statistic for a comparison of two diagnostic procedures is calculated as the ratio of $(K_1 - K_2)$ and the standard deviation of $(K_1 - K_2)$. The constant λ is eliminated in this ratio.

Announcement

The **Society for Health Services Research in Radiology** (SHSRR) has recently been chartered with the goals of promoting health services research and education in radiology through (1) establishing educational programs; (2) developing forums in which investigators can present their research; (3) fostering collaborations in research among its members; (4) encouraging the development of careers in health services research; and (5) disseminating members' research findings.

The SHSRR will hold annual meetings comprised of short instructional courses, a postgraduate educational program, and presentations of competitive abstracts proffered by both members and nonmembers. The first annual SHSRR meeting will be held in Philadelphia, PA, September 9-11, 1996, in cooperation with the third annual postgraduate course on health services organized by the University of Pennsylvania.

Membership in the SHSRR is open to both physicians and nonphysicians with an interest in technology assessment, decision analysis, health financing and economics, quality assurance, epidemiology, biostatistics, research methods development, informatics, disease management, or related fields. Individuals interested in joining the SHSRR should contact Bridgette Bienacker, Executive Director, SHSRR, c/o American College of Radiology, 1891 Preston White Drive, Reston, VA 22091.